Gerald Amiel Ballena

✉ gmballena@up.edu.ph   🐙 github.com/GABallena   in linkedin.com/in/gerald-amiel-ballena

📍 Metro Manila, Philippines   🆔 orcid.org/0009-0000-8857-9755

Professional Summary

Results-driven Bioinformatics Specialist with 3+ years of experience in NGS data analysis, pipeline development, and computational biology. Proven expertise in reducing processing time by 75% and managing 2.5TB+ daily genomic data workflows. Skilled in Python, R, and high-performance computing for environmental metagenomics and machine learning applications.

Skills

- Next-Generation Sequencing (NGS)
- Python Programming
- R Statistical Computing
- Machine Learning & AI
- Pipeline Automation
- High-Performance Computing (HPC)
- Data Analysis & Statistics
- Metagenomics Analysis
- Docker Containerization
- Git Version Control
- SQL Database Management
- Bioinformatics Tools

Professional Experience

Project Technical Specialist: 2024 – Present

University of the Philippines, College of Public Health

- Engineered 15+ production-grade bioinformatics pipelines processing 2.5TB+ genomic data daily
- Optimized computational workflows achieving 75% reduction in resource utilization
- Implemented parallel processing architecture handling 1000+ concurrent metagenomic samples
- Reduced analysis time from 6 hours to 30 minutes through automated workflow optimization
- Developed statistical frameworks for environmental monitoring with 99.9% accuracy

AI Biology Specialist: 2024 – Present

Outlier.ai (Remote)

- Provided biology expertise to train generative AI models, ensuring accuracy and relevance in domain-specific outputs.
- Assessed, reviewed, and ranked AI-generated text for factuality, clarity, and scientific integrity.
- Crafted and evaluated STEM-related questions to refine AI responses and improve user satisfaction.

Technical Skills

Programming & Development:

- **Programming Languages:** Python, R, Bash, SQL
- **Bioinformatics Tools:** BioPython, Bioconductor, BLAST, BWA, SAMtools, Kraken2, MetaPhlAn, MEGAHIT, SPAdes, BUSCO, QUAST
- **Workflow Management:** Snakemake, Nextflow, Docker, Conda

References available only upon request and if shortlisted

- **Version Control:** Git, GitHub, GitLab
- **Databases:** PostgreSQL, MongoDB, CARD, RefSeq

Key Projects

Environmental Metagenomics Pipeline:
- Developed end-to-end workflow for taxonomic and functional profiling

Statistical Analysis Framework:
- Created automated system for environmental data analysis
- Implemented advanced statistical methods for contamination detection
- Built visualization tools for quality metrics and diversity indices

Education

University of the Philippines Diliman

Graduated: July 2022

Thesis: In silico assessment of the association of pathogenicity and metal-resistance potential of *Fusarium* spp. Pre-print Link

Accomplishments: DOST ASTHRDP-Scholarship
Calculated Effective GPA (MS): 1.72

Relevant Courses

- Industrial Microbiology
- Molecular Phylogenetics (Maximum Likelihood, Bayesian)
- Epigenetics (Chromatin Remodeling, Histone Modifications)
- Advanced Cell and Molecular Biology (Gene Expression, Signal Transduction)
- Microbial Physiology and Metabolism

University of the Philippines Baguio

Graduated: June 2018

Thesis: **Bioelectrocatalysis by Novel Electrogenic Alkaliphilic Bacteria *Bacillus* sp. BAB-3442 Using Dual-Chambered Microbial Fuel Cell** .... Poster Presentation (PSM 47)

Accomplishments: Advanced Placement Exam: Advanced Algebra

Laboratory and Technical Skills

Advanced Laboratory Techniques:
- Molecular Biology: PCR, Western Blotting, DNA/RNA extraction, Bacterial transformation
- Microbiology: Bacterial isolation, Antibiotic susceptibility testing, Strain development
- Bioinformatics: Phylogenetic analysis (Maximum Parsimony, Bayesian Inference, Maximum Likelihood)
- Cell Biology: Stem cell research, Histological techniques, Animal handling
- Industrial Processes: Ethanol distillation, Fermentation optimization

Core Competencies:
- Experimental design and protocol optimization
- Quality control and documentation
- Laboratory safety and compliance

References available only upon request and if shortlisted

- Scientific paper analysis and critique

Certifications and Relevant Coursework

Certifications

Data Analyst Associate: DataCamp 2024

Data Engineer Associate: DataCamp 2024

AI Fundamentals: DataCamp 2024

Data Literacy: DataCamp 2024

Principles, Statistical and Computational Tools for Reproducible Data Science: Harvard EdX . . 2024

Professional Development and Online Certifications

Data Science & Programming:
  - **Python Specialization** (DataCamp, 2024)
    - Data Analysis, Machine Learning, Statistical Analysis
    - Scikit-learn, Pandas, NumPy, Matplotlib
    - Geospatial Analysis and Visualization
  - **Data Engineering Certification** (DataCamp, 2024)
  - **R Programming Specialization** (DataCamp, 2024)
    - Advanced Data Visualization
    - Statistical Analysis and Modeling

Bioinformatics & Genomics:
  - **Genomic Data Science Specialization** (Coursera, 2024)
    - Bioconductor, RNA-Seq Analysis, CHIP-Seq
    - Python and R for Genomic Analysis
    - Command-line Tools for Genomics
  - **Advanced Bioinformatics** (Coursera, 2024)

Public Health & Biostatistics:
  - **Biostatistics in Public Health** (2024)
    - Statistical Analysis and Hypothesis Testing
    - Regression Analysis and Epidemiologic Tools

Key Achievements

  - Built and deployed metagenomics pipeline currently processing >2.5TB genomic data annually
  - Decreased pipeline runtime through workflow optimization
  - Automated data processing workflows reducing analysis time per dataset

Selected Projects

Environmental Data Analysis Platform:
  - Developed comprehensive R/Python pipeline for physicochemical data analysis
  - Implemented statistical modeling with multiple regression, PLS, and machine learning
  - Created interactive visualizations for water quality parameters and sampling sites

Geospatial Analysis Framework:
  - Built automated mapping system for environmental sampling locations
  - Integrated multiple data sources including OpenStreetMap and local databases
  - Generated publication-quality maps with statistical overlays

Metagenomic Analysis Pipeline:
  - Developed pipeline capable of processing large-scale genomic data with multi-threaded processing
  - Successfully analyzed 26 metagenomic samples
  - Implemented parallel processing optimization, enabling efficient batch analysis across multiple CPU cores

Metagenomic Pipeline Development:
  - Developed end-to-end workflow for taxonomic profiling and diversity analysis
  - Integrated Kraken2, Bracken, and scikit-bio for

References available only upon request and if shortlisted

automated processing

- Standardized pipeline outputs across multiple datasets through version control and containerization

Statistical Optimization Framework:
- Created automated workflow for k-mer profile analysis
- Implemented statistical model selection and validation
- Reduced analysis time from days to hours

Statistical Analysis Automation:
- Designed reproducible workflows for multivariate environmental data analysis
- Implemented automated quality control and outlier detection systems
- Created standardized reporting templates for research findings
- Built parameter optimization system for multiple trimming tools (Trimmomatic, fastp, BBDuk)

Automation Systems:
- **Version Management:** Developed automated system for tracking and updating Bioconda dependencies
- **Bootstrapping Framework:** Created reproducible analysis system with automated parameter exploration
- **Workflow Management:** Built scalable pipeline for processing 1000+ metagenomic samples
- **Resource Optimization:** Implemented parallel processing and memory management for HPC environments